

Knowing What We Know about IT and Business Value: Cause for Concern about Endogeneity Problems and Potential Solutions

M.Lynne Markus
Bentley University, USA
MLMarkus@Bentley.edu

Frantz Rowe
LEMNA, Nantes University and
KTO, SKEMA Business School, France
frantz.rowe@univ-nantes.fr

Abstract

Do IT investments deliver business value? This long-standing question of IS interest is a causal question. Answers to this question are often sought through the use of econometric methods, which require careful attention to the issue of endogeneity for valid causal inference. Yet, concerns about endogeneity problems in econometric research persist despite the many quantitative techniques available for addressing them. Recent publications in strategic management and accounting have offered a few non-quantitative solutions, such as better writing and reviewing norms, better theory selection, and use of descriptive quantitative and qualitative methods. Not considered in these prescriptions is a relatively little-known category of explicitly causal case study research methods that originated in sociology and political science. This paper describes these methods, shows how they address endogeneity problems, and explores how they might complement statistical methods in the study of IT business value.

1. Introduction

The contribution of IT to business value has been and continues to one of the most important topics of Information Systems research, the subject of hundreds of quantitative and qualitative papers over the years. Although the consensus of the IS field is that IT certainly matters for business value, “the fundamental question of the causal relationship ... remains partly unexplained” [36] (p. 1).

Many theoretical and empirical factors undoubtedly contribute to this state of affairs, including conceptual fuzziness and measurement problems [36]. An additional factor that has, however, received little systematic attention in the IS literature is the problem of endogeneity in the quantitative models that are a primary approach for investigating IT’s business value.

Endogeneity is usually defined in *statistical* terms and can be caused by data scarcity and measurement error. But it is sometimes caused by *theoretical* problems such as the omission of causal variables from the models tested or the possibility of reverse causality or feedback loops.

Endogeneity is important for IT business value research because it casts doubt on the validity of the research findings. Consider, for example, a recent econometric study aiming to estimate the contribution of big data and analysis to firm performance [30]. Well aware of the threat endogeneity poses for causal inference, Müller et al. [30] introduced instrumental variables into their analysis “to control for potential biases arising from reverse causality or omitted variables” (p. 501). Despite these precautions, however, they found that “the direction of causality between [big data and analytics] and firm performance in these models remains unclear” (p. 502).

Although endogeneity has been explicitly addressed in individual IT business value studies (cf. [30] [3] [1] [2]), it has not been explored systematically in the IT business value literature as a theoretical issue. This is a sad gap in our literature. It is surely essential to the IS body of knowledge to have a compelling answer to question of whether IT investments improve firm performance or whether it is that mainly high performing firms invest in IT. It is equally important to explain “the unexplained process of creating internal and competitive [business] value” [36] (p. 31). Accordingly, it is the aim of this paper to discuss endogeneity as a *theoretical* problem in IT business value research and to describe a class of empirical methods that may complement statistical approaches by suggesting better theories.

In the second section of this paper, we review recent discussions about the problem of endogeneity in various management-related fields, including IS. We explain that endogeneity is not just a statistical problem that can be addressed by statistical remedies, but also a theoretical problem. To illustrate that point, we give examples of three important endogeneity

problems in the IT business value literature. In the third section, we outline the non-statistical remedies that have been proposed in various management literatures for dealing with theoretical endogeneity. Regrettably, these remedies do not include an entire body of empirical research methods that could provide evidence useful for improving and providing convincing justification for quantitative models of IT business value. We offer a brief overview of these methods and give examples of how they could be used to address theoretical endogeneity in IT business value research. We conclude with a discussion of how to forge a workable partnership between quantitative and qualitative causal methods in Information Systems.

2. Endogeneity Problems in Management and IT Business Value Research: From a Statistical Problem to a Theoretical Problem

Management disciplines such as accounting, marketing, strategic management, and IS have long relied on multivariate methods of statistical association for testing their theories. The universally acknowledged disclaimer that “correlation does not equal causation” did not carry much weight historically [39] in the face of epistemological beliefs that causality is an unobservable phenomenon, or even a metaphysical concept [28], especially in the absence of methods designed to support causal analysis. The past quarter century or so, however, has seen a “causal revolution” [33] in which the identification of causal relationships has become a key goal and in which both quantitative [33] and qualitative [34] [4] tools have been developed for causal inference.

Attending this revolution in scientific aims and tools has been a growing attention to the problem of endogeneity in management disciplines [18] [12] [16] [37] [43] [20]. Endogeneity is often defined in entirely statistical terms as a situation in which an independent variable is correlated with the error term in an ordinary least squares (OLS) regression and is often said to arise because of data scarcity and measurement errors [37] [20] [43]. When the problem is described this way, proposed solutions to the problem include a long list of statistical techniques, including instrumental variables approaches with two- or three-stage least squares estimation, non-instrumental approaches such as higher moments, identification through heteroscedasticity, the Heckman method, latent variables or coupla, natural or quasi experiments, lagging independent variables, and regression discontinuity designs [43] [38]. These

approaches are well represented in the IS literature [13, 23]

Despite the availability of many quantitative techniques for dealing with endogeneity, they are often difficult to apply. For example, Gow et al. claim that “the assumptions required to apply quasi-experimental methods [e.g., the difference-in-differences approach] are unlikely to be satisfied by observational data in most empirical accounting research settings.” [16] (p. 479) “[N]o statistical test allows the researcher to verify that their instruments satisfy [criteria for the validity of their instruments]” [16] (p. 489). Similarly, Ketokivi and McIntosh claim that so-called exogenous variables “*always* are [endogenous], least to an extent” [20] (p. 4, original emphasis). But “to expect perfect exogeneity would set an unreasonable standard for empirical research.” [20] (p. 3) “Instead,” they say, “we need to determine whether the problem is enough of a cause for concern in that it materially jeopardizes estimation” [20] (p. 4).

Although it can be difficult or impossible to completely rule out endogeneity, “reviewers and editors in multiple disciplines have increasingly identified endogeneity as an *alternative explanation* for results presented in papers they evaluate, and endogeneity represents a more and more frequent reason for manuscript rejection” [37] (p. 1070, emphasis added). Commenters such as [20, 38] argue that using endogeneity as a “fatal flaw” rationale for rejecting papers is an inappropriate norm for management research. A more appropriate stance is to require authors to justify their models theoretically:

“[I]t is reasonable to ask that authors consider well-known or obvious alternative explanations if they wish to advance a new claim. It is unreasonable to ask them to prove there cannot be any alternative explanation.” [38] (p. 10)

In short, researchers in a number of management fields believe endogeneity to be a serious problem that cannot be fully remedied by statistical techniques. Furthermore, it is not merely a statistical problem but also a theoretical one, requiring attention to alternative explanations. In other words, addressing theoretical endogeneity is not just a matter of finding relevant variables that are uncorrelated with a model’s error term, but rather a matter of explaining why certain variables are, or are not, hypothesized to be causal.

Thinking of endogeneity as an *alternative explanation* for empirical results is a much broader interpretation than thinking of endogeneity as a statistical problem. The management literature on endogeneity mentions two alternative explanations that might plague IS research on IT business value—self-selection and simultaneous or reverse causality. We also consider a third type of theoretical endogeneity—

complements and synergies—which has been important in IT business value research. We consider these challenges to be *theoretical* endogeneity problems, as opposed to merely statistical endogeneity problems arising from poor choice of variables or measurement error.

2.1 Self-Selection.

“[W]e generally think that managers’ desire to achieve high levels of performance influences their decisions about [which strategic decisions to make]. If this presumption is correct, then managers make strategic organizational decisions not randomly, but based on expectations of how their choices affect future performance. Put more precisely, the field of strategic management is predicted on the idea that management’s decisions are endogenous to their expected performance outcomes—if not, managerial decision-making is not strategic; it is superfluous.” [18] (p. 51)

Rephrased, this means that managers’ decisions to make investments in ITs such as big data and analytics are based on expectations of how much they might benefit from those investments. Not all organizations are equally well placed to benefit from investments in big data and analytics. Data analytics can make use of internal (e.g., enterprise transactional) or external (e.g., social media) data or both. According to big data experts, high quality data is essential for successful analytics use. This suggests the hypothesis that among those organizations aiming for customer service innovation (for example see [25]) organizations that have already invested heavily in high quality internal customer data, e.g., through ERP and CRM, are more likely to be able to benefit from investments in big data and analytics than organizations that have not done so. Because they are more likely to be able to benefit, it follows they are more likely to invest, that is, to self-select into the big data and analytics investment condition.

Self-selection is an important theoretical case of endogeneity caused by omitted variables [12], and it is not easily controlled by the means employed in the Müller et al. study (introducing the presence of ERP and CRM systems into their equation as control variables), because that approach does not take into account *the quality of the customer data* achieved (or not) through investments in enterprise systems.

2.2 Simultaneous or reverse causality.

A second type of theoretical endogeneity concerns what is often called simultaneous or reverse causality. An earlier mentioned example of the reverse causality

problem was whether firms that invest in IT have higher performance than those that do not, or whether higher performing firms invest more in IT than lower performing firms. This was the theoretical challenge that the Müller et al. study attempted to avoid with limited success.

Another example draws on the theoretical synthesis [17] on strategic business value from big data analytics (BDA) that headed up the special issue in which the Müller et al. paper appeared. Grover et al. characterized the value creating process as consisting of two phases, BDA capability building and BDA capability realization, with a feedback loop labeled “learning by doing” arcing from capability realization to capability building. Consider how this model could be causally modeled for quantitative analysis. A basic assumption is that the greater the capabilities created through investments in BDA, the greater the benefits received. Since investments occur before benefits can be received, a researcher might build a model that considers the size of investment at time t_1 to be an exogenous cause of the size of the benefits at time t_2 .

This theoretical model, however, does not fit our collective knowledge about how firms build their IT investments. Aware no doubt of IS field writings on real options theory [5], organizations often approve initial investments in unproven technologies and then condition their future investments on benefit realization outcomes. In other words, we have good reason to believe that for many firms, there actually is a feedback loop between benefit received and BDA investment, as Grover et al. [17] point out; therefore, investment is not an exogenous variable [2].

2.3 Complements and synergies.

A third type of theoretical problem observed in the IT and business value research has not been discussed in the management literature as a type of endogeneity per se, but undoubtedly has implications for quantitative modeling. This is the problem of complements and synergies.

It has long been recognized in the IS literature that that getting business value from information technology depends on coordinated investments in other things, including training (of users and technical staff), infrastructure and data, other technologies, business process changes, and the like. Brynjolfsson [7] developed this argument carefully in his 2000 paper “Beyond Computation” and with Aral [1] in 2012 provided a convincing econometric demonstration of three-way complementarities between ERP systems, HR analytics, and pay-for-performance compensation schemes.

Complementarities and synergies can be identified through regression-type research, if the problem is theoretically modeled that way. The theoretical endogeneity problem comes in when models are based on the assumption that IT makes an independent, rather than an interdependent, contribution to business value. As Schryen observed:

“[T]he analysis of the impact of IS on selected outputs suffers the problem that firm performance is also *simultaneously* affected by a host of other internal and external factors, making it exceedingly difficult to isolate *the influence of IT alone...*” [36] (p. 11, emphasis added)

Returning to the Müller et al. paper, the authors noted that their “results can probably be explained by the fact that BDA solutions require complementary IT assets and capabilities, such as transactional enterprise systems or data scientists, which can provide the necessary data and skills to extract knowledge out of this data” [30] (p. 504). Interestingly, while they had data on the use of ERP, CRM, and SCM systems (and used this data in their model for control and instrumental variables), they did not add the interaction terms that would have allowed them to test for complements.

Other IT business value researchers including [29, 42] have attempted to assess complements and synergies using the qualitative case comparative method, fs/QCA. This is one of a family of methods, which we consider below, that includes the single case study. Single case studies have also been used to explore complements and synergies. For instance, in a qualitative case study of Grupo Santander’s phenomenal growth over the last 25 years, Parada et al. [32] showed that the company systematically invested in human resources while deploying its integrated banking systems in acquired companies. Additional synergies are revealed in more complex cases [9].

Self-selection, simultaneous or reverse causality, and complements and synergies are three theoretical problems plaguing IT business value research that cannot always be convincingly addressed with the techniques proposed for coping with statistical endogeneity. There are undoubtedly other theoretical endogeneity problems. In the next section, we consider various solutions for theoretical endogeneity problems.

3. Remedies for Theoretical Endogeneity

Management literatures exhibit awareness that theoretical endogeneity (plausible alternative explanations not ruled out through modelling) calls for solutions that go beyond statistical solutions like the

use of instrumental variables and the Heckman approach. We found at least three alternative solutions in the management literatures: adopting better social norms for writing and reviewing papers, using better causal theory, and relying on deep description.

3.1 Social norms.

For Shaver, the systemic solution is to instill norms of writing and reviewing to build a cumulative body of research that sheds light on causal questions:

“We recognize that any one paper will not solve the identification challenge for most strategy and organization questions. Rather, it will take multiple papers that in combination will provide insight into the causal mechanisms.” [38] (p. 3)

Among Shaver’s recommendations for moving toward a cumulative body of research are: “denounce novel theory as a publication requirement” [38] (p. 9) and “include a detailed discussion of the research design strategy and the assumptions under which the results will reflect causal effects” [38] (p. 10).

While the remedy of better reviewing norms may help researchers avoid summary rejection, it does not address the fundamental issue of theoretical endogeneity—the need for sound and plausible theories to undergird mathematical models.

3.2 Better theory.

For accountants Larcker and Rusticus [24], the systemic solution to endogeneity problems in research about firm ownership and performance is to use better theory:

“Since empirical studies typically work with a cross-section of firms, it is necessary to assume that the relation between ownership and performance is similar for all firms It is hard to imagine a situation [consistent with these assumptions], and this suggests that the above assumptions are not appropriate for this kind of empirical study.” [23] (p. 209) ... “Without a theory of ownership choice it is not obvious which variables are exogenous and should be include in the prediction model, and which are endogenous and should be excluded.” [24] (p. 213) ... “We believe that using a one-size-fits-all approach and treating these variables as exogeneous is seriously flawed, and should not be guiding future research. ... [I]t will be necessary to have much better theory guiding the development of empirical studies and the choice of exogeneous variables.” [24] (p. 214)

Similarly, accountants Gow et al. [16] call for “clear discussion” about the causal mechanisms hypothesized in causal models.

Although the value of better theory seems unquestionable, we would be remiss not to include a contrary viewpoint. van Lent [41] explicitly disagreed with Larcker and Rusticus [24] on the grounds that “theory is often non-existent or its assumptions too strong to admit application” [41] (p. 198). Accordingly, van Lent argued that “theory is just not specific enough to ensure whether endogeneity is present or absent in a given context and I doubt whether theory will ever achieve a level of development in which it will admit unequivocal judgement on this question.” [41] (p. 201)

van Lent’s argument seems to depend on a particular understanding of theory, of the “if more X, then more Y” sort. However, there are many different types of theoretical statements [28], and theoretical statements of a type different from that envisioned by van Lent might be useful for generating alternative explanations or suggesting variables for inclusion in IT-business value research models. Furthermore, the alternative explanations needed to address theoretical endogeneity can be discovered by means of qualitative empirical methods. Modelers may be able to improve their work by drawing on qualitative research conducted by other scholars.

3.3 Deep description.

A third proposal for solving the problem of theoretical endogeneity is deep description. Gow et al. explain their suggestion as follows:

“Accounting is an applied discipline and it would seem that most empirical research studies should be solidly grounded in the details of how institutions operate. Unfortunately, there are very few studies published in top accounting journals that focus on providing deep description of institutions relevant to accounting research settings. ... One reason to value descriptive research is that it can uncover realistic structures and mechanisms that would be exceedingly difficult to arrive at from basic econometric theory.” [16] (p. 499)

Gow et al. [16] cite examples of descriptive studies, employing both quantitative and qualitative data, that offer insights that could be useful to causal modelers. But there is nothing in their account of deep description that shows an awareness of extant qualitative methods have been expressly designed for the development or testing of explicitly causal theory. Below, we briefly describe the class of methods called causal case study methods. Subsequently, we show how they might be used to address theoretical endogeneity in IT-business value research.

3.4 Causal case study methods.

According to Tsang [40], there are at least four different ways of theorizing from case studies: 1) identification of empirical regularities, 2) theory building and testing, 3) interpretive sensemaking, and 4) contextualized explanation. Causal case study methods [15] [4] address both contextualized explanation (in a single case) and theory building and testing (across a number of cases).

In parallel with the “causal revolution” in quantitative methods described by Pearl and Mackenzie [33], impressive developments have occurred in qualitative methods explicitly aimed at posing and answering causal questions. To the best of our knowledge, these methods originated in the fields of sociology and political science, but they are slowly being taken up in several management disciplines, including IS. The best known of these techniques is Qualitative Comparative Analysis (QCA) and its fuzzy-set complement (fs/QCA) which, based on Boolean algebra, analyze the extent to which characteristics or conditions are present or absent in a set of observations [35]. A related method is NCA (necessary condition analysis) [14].

QCA, fs/QCA, and NCA are similar to regression-type methods in that they are cross-sectional or comparative, but the family of causal case study methods also include in depth analysis of single cases. In the analysis of single cases, the key technique is known as process tracing, which consists in the identification and conceptualization of sequences of events in time. Thus, it can be very useful in sorting out questions related to self-selection and recursion or reverse causality.

In general, causal case study methods differ from regression-type analysis in several important ways. First, causal case study methods involve a completely different *research objective* than econometric methods [28]. According to Pearl and Mackenzie, the goal of quantitative causal analysis is to apply *existing causal reasoning* to empirical data in search of practical answers:

“Many people still make [the] mistakes of thinking that the goal of causal analysis is to prove that X is a cause of Y or else to find the cause of Y from scratch. That is the problem of causal discovery, which was my ambitious dream when I first plunged into graphical modeling In contrast, the focus of ... this book, is representing plausible causal knowledge, combining it with empirical data, and answering causal queries that are of practical value.” [33] (pp. 79-80)

By contrast, causal case study methods aim precisely to *discover* potential causes or to *provide convincing evidence* that causal connections exist. According to Kreuzer, causal process tracing, a key causal case study method,

... “has a deductive and a theory-testing orientation that focuses on the observable implications of hypothesised causal mechanisms within a case to test whether a theory on these mechanisms explains the case Second, process tracing also has an ‘inductive, theory-developing side that uses evidence from within a case to develop hypotheses that might explain the case’ ...” [21] (p. 473)

A second key difference between causal case study and econometric methods is a completely different understanding of *causality* [28]. Statistical analysis assumes that variables—elements, features, or factors liable to vary or change—are causes that raise the probability that outcomes will be observed within a population of entities [27]. One looks for causal effects by analyzing a population of entities exhibiting variation in causes and effects and attempts to isolate the effects attributable solely to the presumed cause or causes. By contrast, causal case study methods understand causality as the operation of causal processes or mechanisms, defined by Mahoney [26] as events and by Beach and Pedersen [4] as a theoretical system of actors performing activities within individual cases. For instance, whereas a researcher using econometric methods might theorize that greater investment in BDA increases business benefits, the causal case study researcher might focus on organizations’ creation of new product offerings from analysis of customer data, which then contribute to higher revenues.

A third key difference is the unit of analysis. In econometric approaches, the unit of analysis is a natural system such as a person, an organization, or a project team. These units may be called cases in common parlance, but there is a very different understanding of the “case” in causal case methods. In causal case study methods, the case is understood as abstract “instance of a causal process playing out, linking a case (or set of causes) with an outcome” [4] (p. 5). For instance, whereas the econometrician may compare BDA outcomes across *organizations*, the causal case study researcher might compare cases of *improved firm performance* involving new product offerings suggested by BDA.

Some authors have attempted to combine qualitative causal case study methods with quantitative causal methods under a unified theoretical framework, cf. [27], but most causal case study experts

differentiate sharply between the two [4]. A single point of routine practice will illustrate the vast differences between these two families of methods for causal analysis, that nevertheless remain highly complementary. Whereas quantitative causal methods aim to examine an entire population (or a representative sample) of entities for analyzing the full range of variation on the dependent variable, qualitative causal researchers often select cases that exhibit a specific outcome, so that they can isolate the causal processes in cases where the outcome of interest actually occurred [15] [4].

Just like econometric methods, causal case study methods are no panacea. fs/QCA in particular has been criticized for confirmation bias [22] and aggregation bias [6]. Similarly, process tracing in single case studies is a “high tariff” method [19] that also raises concerns about representativeness. However, these specific criticisms do not apply to the entire causal case study family, and they do not diminish the ability of methods such as NCA, QCA, and within-case process tracing to shed light on problems of theoretical endogeneity:

“As Kreuzer (2016) makes clear, Bayesian process tracing can provide a useful framework for discussion whether a given explanatory factor is biased by endogeneity effects. He shows how evidence concerning the intervening events that link X [presumed cause] and M [mechanism] can help rule out endogeneity. More generally, I would emphasise that hypotheses about the causes of specific outcomes in particular cases—which is the focus of causal inference in qualitative process tracing—do not face identification problems in the same way as observational research concerned with estimating the average effects of variables. ... The evaluation of hypotheses [about the causes of specific outcomes in particular cases] does not require the use of control variables in order to establish causality.” [26] (p. 497)

3.5 Causal case study methods as remedies for theoretical endogeneity.

We now return to the three types of theoretical endogeneity in IT business value research that we presented earlier, aiming to show how researchers using causal case study methods might seek to address them. Naturally, as with quantitative analysis, any number of causal case study research designs could possibly address these issues. We aim only to illustrate how it might be done.

Self-Selection. How might causal case researchers approach the problem of deciding whether there is a problem of self-selection in BDA investments? One approach would be carefully select cases for study on theoretical, not just on practical, grounds. For instance, in the domain of service innovation [25], one might select and focus on cases of successful service innovation (the outcome) AND internal customer data, whether or not BDA was used. (The qualifier “AND internal customer data” would be needed to rule out the use of packaged solutions provided by external providers relying on external social media data.) Naturally, doing this well would involve considerable effort in defining service innovation (e.g., deciding whether to include in the definition both automated and human-material customer service inter/actions as in the Lehrer et al. typology) and in screening cases to see whether they actually fit the definition.

Then the researcher might trace the pathways to service innovation in the cases, either trying to surface theory or to test hypotheses like: Is BDA a necessary condition for service innovation success? Or, is BDA plus ERP and CRM sufficient for service innovation success? If ERP and CRM were found to be unnecessary for cases in which BDA contributed causally to success, did non-users of ERP or CRM experience greater difficulties or employ different strategies to bring out success? Did initial non-users of ERP or CRM find that they had to acquire good transactional system data before they could proceed with BDA? Answering questions like these, even in a handful of cases, could be enough evidence to provide plausible answers to whether self-selection would create endogeneity problems in a quantitative analysis.

Recursive causality. Causal case study methods can also be used to reduce concerns over simultaneous or reverse causation. Büthe [10] argues persuasively that one need not worry about the endogeneity of a cause if one can show through temporal analysis that once an outcome is observed, the hypothesized cause can be seen to amplify (or dampen) the effect. Naturally, this kind of analysis is very challenging to do quantitatively in the absence of superb time series data. But it is a relatively straightforward causal case study exercise, because with process tracing one can learn whether BDA preceded and influenced performance outcomes, the other way around, or both. Qualitative process tracing can provide the kind of evidence that would lend plausibility to (or cast more doubt on) the theoretical assumptions underlying a particular econometric model.

Conjunctive causality (complements and synergies). Finally, demonstrating (or debunking) complementarities and synergies is perhaps the easiest proof of concept for comparative causal case study methods like QCA or fs/QCA. And in fact, as mentioned earlier, fs/QCA has been used in IS literature research to explore the possibility of complements and synergies in IT business value research [29, 42]. Practitioners of comparative causal case study assume that causality is almost always complex, that is, that causes in social settings rarely, if ever, operate alone, but usually in conjunction with other causes. The task is then to hypothesize the other factors that would work in conjunction with BDA and to include them in an analysis of an intermediate number of cases (ideally, according to Ragin [34], with variation on the outcome). This type of analysis would yield up a reduced number of causal configurations associated with the outcomes of interest, which could then be analyzed in a quantitative population-type study.

3.6 Summary.

Theoretical endogeneity in management and IT business value research is a bigger challenge than statistical endogeneity, because it can remain even when independent variables are uncorrelated with error terms in regression analyses. Purely statistical techniques will not resolve theoretical endogeneity. (See Table 1 for a comparison of statistical and theoretical endogeneity.)

Various approaches for dealing with theoretical endogeneity have been discussed in the management literature, including reviewing norms (not expecting a solution to theoretical endogeneity), better theory, and deep description. To those solutions, we add the family of causal case methods, which include both single-case approaches like process tracing, and cross-case approaches like QCA and NCA.

Naturally, we do not propose that econometricians adopt causal case study approaches as substitutions for statistical methods. However, we do believe that econometric modelers can gain inspiration for better models (ones that avoid theoretical, as well as statistical, endogeneity) by reading and citing research that employs causal case approaches.

Table 1. Statistical Endogeneity Versus Theoretical Endogeneity

	Statistical Endogeneity	Theoretical Endogeneity
Definition	Occurs when an independent variable is correlated with an error term in a regression model	Occurs when a theoretical model (quantitative or qualitative) is incorrect
Causes	<ul style="list-style-type: none"> • Data scarcity • Measurement error 	<ul style="list-style-type: none"> • Self-selection • Simultaneous or reverse causality • Complements and synergies
Remedies	<ul style="list-style-type: none"> • Instrumental variables approaches with two- or three-stage least squares estimation • The Heckman method • Latent variables or coupla • Lagging independent variables • Regression discontinuity designs 	<ul style="list-style-type: none"> • Better editorial norms • Better theory • Deep description • Casual case study methods: <ul style="list-style-type: none"> ◦ Within case process tracing ◦ Across case QCA and NCA
Comments	Statistical remedies do not address theoretical problems that arise from an unsound explanation of the hypothesized relations between causes and effects	Causal case study methods are qualitative and mathematical empirical approaches that can support better theoretical arguments, thereby complementing statistical techniques

We should point out that quantitative modelers might also find similar inspiration from descriptive research (quantitative or qualitative) as suggested by Gow et al. [16]. We include in this category Tsang's [40] interpretive sensemaking and identification of empirical regularities types of theorizing. But we also believe that causal case study methods have a distinct advantage over descriptive research in that the former methods are designed, with acute attention to the challenges of causal inference, expressly for the purpose of answering causal questions. Most descriptive research methods are not explicitly causal.

4. Discussion

The contribution of IT to business value is usually studied with quantitative causal methods. In this paper we examined an alternative — qualitative — family of research methods designed to answer explicitly causal questions. Although some scholars have asserted that “process theory” and “variance theory” can profitably be combined in empirical research [31] [8], quantitative and qualitative *causal* methods are not easily combinable [4], because they differ significantly in their philosophical assumptions. Furthermore, these two families of methods differ substantially in their skill

requirements: Individual researchers or small research teams are unlikely to excel at both.

Nevertheless, quantitative and qualitative causal methods are highly complementary, because each produces knowledge that can be useful in addressing the limitations of the other. Whatever the type of empirical investigation, quantitative or qualitative, causal inferences always depend on access to the data and to the field [4]. While exceptional quantitative datasets may help to rule out endogeneity problems, e.g. [1], econometric techniques are typically limited by the datasets to which they can be applied. Single causal case studies can help fill in the gaps created by lack of quantitative data.

On the other hand, process tracing in single case studies is a “high tariff” method [19] that also raises concerns about representativeness. Despite these concerns, case studies can provide unique insights [11] into causal processes. We believe that quantitative modelers can benefit from qualitative insights to help them decide how to model causal questions, which data to seek, which variables to choose when choice is an option, and how to justify their research design decisions. In other words, these methods can provide what Pearl [33] called “plausible causal knowledge” for use in quantitative causal models. Similarly, we believe that qualitative researchers can benefit from quantitative insights to help them decide where to spend their scarce process tracing resources and what to look for in case studies.

5. Conclusion

Recent discussions about the problem of endogeneity in various management-related fields and in the social sciences offer strategies for clarifying causal thinking and enhancing the validity of causal inferences. The management literatures propose adopting better social norms for writing and reviewing papers, using better causal theory, and relying on deep description as alternatives to well-known quantitative approaches for dealing with endogeneity problems. Qualitative causal case study research methods originating in sociology and political science can also be beneficial for progress in IT business value research.

However, both quantitative and qualitative researchers would have to give something up in order to benefit from the synergies between their preferred causal methods and effectively address theoretical endogeneity problems. Quantitative researchers would have to give up the hope that all causal questions worth answering can be answered with quantitative methods. They would have to accept that qualitative causal methods offer valuable insights into causal *discovery* [33] as well as hypothesis testing [4] [21]. Qualitative researchers would have to give up the belief that “quantitative” and “causal” are anathema in the study of human affairs. They would have to embrace quantitative findings as clues about where to focus their high tariff research strategies on high payoff research questions. But, both camps, we believe, would gain much of value. And the study of IT business value would come out way ahead.

6. References

- [1] Aral, S., E. Brynjolfsson, and L. Wu, “Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology”. *Management Science*, 2012. 58(5): p. 913-931.
- [2] Baker, J., J. Song, and D.R. Jones, “Closing the Loop: Empirical Evidence for a Positive Feedback Model of IT Business Value Creation”. *The Journal of Strategic Information Systems*, 2017. 26(2): p. 142-160.
- [3] Barua, A. and B. Lee, “The Information Technology Productivity Paradox Revisited: A Theoretical and Empirical Investigation in the Manufacturing Sector”. *The International Journal of Flexible Manufacturing Systems*, 1997. 9(2): p. 145-166.
- [4] Beach, D. and R.B. Pedersen, *Causal Case Study Methods*. 2016, Ann Arbor: University of Michigan Press.
- [5] Benaroch, M. and R. Kauffman, “A Case for Using Real Options Pricing Analysis to Evaluate Information Technology Project Investments”. *Journal of Management Information Systems*, 1999. 10 (1): p. 70-86.
- [6] Braumoeller, B.F., “Aggregation Bias and the Analysis of Necessary and Sufficient Conditions in fsQCA”. *Sociological Methods & Research*, 2017. 46(2): p. 242-251.
- [7] Brynjolfsson, E. and L. Hitt, “Beyond Computation: Information Technology, Organizational Transformation and Business Performance”. *Journal of Economic Perspectives*, 2000. 14(4): p. 23-48.
- [8] Burton-Jones, A., E. Monod, and E. McLean, “Theoretical Perspectives in IS Research: From Variance and Process to Conceptual Latitude and Conceptual Fit”. *European Journal of Information Systems*, 2015. 24(6): p. 664-679.
- [9] Busquets, J., “Discovery Paths: Exploring Emergence and IT Evolutionary Design in Cross-Border M&As. Analysing Grupo Santander’s Acquisition of Abbey (2004–2009)”. *European Journal of Information Systems*, 2015. 24(2): p. 178-201.
- [10] Büthe, T., “Taking Temporality Seriously: Modeling History and the Use of Narratives as Evidence”. *The American Political Science Review*, 2002. 96(3): p. 481-493.
- [11] Campbell, D., ““Degrees of Freedom” and the Case Study. *Comparative Political Studies*”, 1975. 8(2): p. 178-193.
- [12] Clougherty, J., T. Duso, and J. Muck, “Correcting for Self-Selection Based Endogeneity in Management Research: Review, Recommendations and Simulations”. *Organizational Research Methods*, 2016. 19(2): p. 286-347.
- [13] Devaraj, S. and R. Kohli, “Performance Impacts of Information Technology: Is Actual Usage the Missing Link?” *Management science*, 2003. 49(3): p. 273-289.
- [14] Dul, J., “Necessary Condition Analysis (NCA) Logic and Methodology of “Necessary But Not Sufficient” Causality”. *Organizational Research Methods*, 2016. 19(1): p. 10-52.
- [15] George, A.L. and A. Bennett, *Case Studies and Theory Development in the Social Sciences*. 2005, Cambridge, MA: MIT Press.
- [16] Gow, I., D.F. Larcker, and P.C. Reiss, “Causal Inference in Accounting Research”. *Journal of Accounting Research*, 2016. 54(2): p. 477-523.
- [17] Grover, V., R.H.L. Chiang, T.-P. Liang, and D. Zhang, “Creating Business Value from Big Data Analytics: A Research Framework”. *Journal of Management Information Systems*, 2018. 35(2): p. 388-423.

- [18] Hamilton, B.H. and J.A. Nickerson, "Correcting for Endogeneity in Strategic Management Research". *Strategic Organization*, 2003. 1(1): p. 51-78.
- [19] Hay, C., "Process Tracing: A Laudable Aim or a High-Tariff Methodology? " *New Political Economy*, 2016. 21(5): p. 500-504.
- [20] Ketokivi, M. and C. McIntosh, "Addressing the Endogeneity Dilemma in Operations Management Research: Theoretical, Empirical and Pragmatic Considerations". *Journal of Operations Management*, 2017. 52: p. 1-14.
- [21] Kreuzer, M., "Assessing Causal Inference Problems with Bayesian Process Tracing: The Economic Effects of Proportional Representation and the Problem of Endogeneity". *New Political Economy*, 2016. 21 (5): p. 473-483.
- [22] Krogslund, C., D.D. Choi, and M. Poertner, "Fuzzy Sets on Shaky Ground: Parameter Sensitivity and Confirmation Bias in fs/QCA". *Political Analysis*, 2015. 23(1): p. 21-41.
- [23] Kumar, N., L. Qiu, and S. Kumar, "Exit, Voice, and Response On Digital Platforms: An Empirical Investigation of Online Management Response Strategies". *Information Systems Research*, 2018. 29(4): p. 849-870.
- [24] Larcker, D. and T. Rusticus, "Endogeneity and Empirical Accounting Research". *European Accounting Review*, 2007. 16(1): p. 207-215.
- [25] Lehrer, C., A. Wienenke, J. Vom Brocke, and S. Seidel, "How Big Data Analytics Enables Service Innovation: Materiality, Affordance, and the Individualization of Service". *Journal of Management Information Systems*, 2018. 35(2): p. 424-460.
- [26] Mahoney, J., "Mechanisms, Bayesianism, and Process Tracing". *New Political Economy*, 2016. 21(5): p. 493-499.
- [27] Mahoney, J., "Toward a Unified Theory of Causality". *Comparative Political Studies*, 2008. 41(4/5): p. 412-436.
- [28] Markus, M.L. and F. Rowe, "Is IT Changing the World? Conceptions of Causality for Information Systems Theorizing". *MIS Quarterly*, 2018. 42(4): p. 1255-1280.
- [29] Mikalef, P., M. Boura, G. Lekakos, and J. Krogstie. "Configurations of Big Data Analytics for Firm Performance: an fsQCA Approach". in *Twenty-fifth Americas Conference on Information Systems*. 2019. Cancun, MX.
- [30] Müller, O., M. Fay, and J. Vom Brocke, The Effect of "Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics". *Journal of Management Information Systems*, 2018. 35(2): p. 488-509.
- [31] Ortiz de Guinea, A. and J. Webster, "Combining Variance and Process in Information Systems Research: Hybrid Approaches". *Information and Organization*, 2017. 27(3): p. 144-162.
- [32] Parada, P., L. Alemany, and M. Planellas, "The Internationalisation of Retail Banking: Banco Santander's Journey Towards Globalisation". *Long Range Planning*, 2009. 42(5/6): p. 654-677.
- [33] Pearl, J. and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*. 2018: Penguin Books.
- [34] Ragin, C.C., *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies 1989*, Berkeley, CA: University of California Press.
- [35] Ragin, C.C., *Fuzzy-Set Social Science*. 2000, Chicago, IL: University of Chicago Press.
- [36] Schryen, G., "Revisiting IS Business Value Research: What We Already Know, What We Still Need To Know, And How We Can Get There". *European Journal of Information Systems*, 2012. 22(2): p. 139-169.
- [37] Semadeni, M., M.C. Withers, and S.T. Certo, "The Perils Of Endogeneity And Instrumental Variables In Strategy Research: Understanding Through Simulations". *Strategic Management Journal*, 2014. 35(7): p. 1070-1079.
- [38] Shaver, J.M., "Causal Identification Through a Cumulative Body of Research in the Study of Strategy and Organizations". *Journal of Management*, 2019 p. forthcoming.
- [39] Shugan, S.M., "Causality, Unintended Consequences and Deduced Shared Causes". *Marketing science*, 2007. 26(6): p. 731-741.
- [40] Tsang, E.W., "Case Study Methodology: Causal Explanation, Contextualization, and Theorizing". *Journal of international management*, 2013. 19(2): p. 195-202.
- [41] Van Lent, L., "Endogeneity in Management Accounting Research: A Comment". *European Accounting Review*, 2007. 16(1): p. 197-205.
- [42] Yi, T.H., S.K. Poon, and J. Chan. "Analysing Causal Complexities in IT Business Value Research". in *Australasian Conferences on Information Systems 2011 Proceedings*. Paper. 2011.
- [43] Zaefarian, G., V. Kadile, H.S. C, and L. Alexander, "Endogeneity Bias in Marketing Research: Problem, Causes and Remedies". *Industrial Marketing Management*, 2018. 65: p. 39-46.